

FROM COPAIR HYPERGRAPHS TO MEDIAN GRAPHS WITH LATENT VERTICES

J.P. BARTHELEMY

Ecole Nationale Supérieure des Télécommunications, 46 rue Barrault, 75634 Paris Cedex 13, France

Received March 11, 1986

Revised October 7, 1987

The purpose of this paper is to extend the Buneman construction of partially labelled trees to the general case. This extension is related with the characterization of median graphs by Mulder and Schrijver.

In the first section, we construct a graph $G(H)$ associated with a copair hypergraph H on a finite set X and define the notion of a median graph with latent vertices (called X -median graph). The latent vertices (i.e. the vertices who are not labelled by elements of X) are obtained by iterating the median operation from actual (labelled) vertices. In the second section, we prove that the graph $G(H)$ is an X -median graph. Then, in the last section, we study some special cases, the Buneman result is reobtained and the hypergraphs whose associated graphs are Hasse diagrams of distributive lattices are characterized.

Résumé. L'objet de ce travail est d'étendre au cas général une construction des arbres partiellement étiquetés proposée par Buneman. Cette généralisation se trouve aussi être fort liée à une caractérisation des graphes médians due à Mulder et Schrijver.

Dans le premier paragraphe, nous construisons un graphe $G(H)$ associé à un hypergraphe complété H (d'ensemble de sommets X) et nous définissons la notion de graphe médian avec des sommets latents (appelé X -graphe médian). Les sommets latents (i.e. les sommets qui ne sont pas étiquetés par les éléments de X) sont obtenus, à partir des sommets "réels" (i.e. étiquetés) en ré-itérant l'opération médiane. Au cours du second paragraphe, on démontre que $G(H)$ est un X -graphe médian. Enfin, dans le dernier paragraphe, le résultat de Buneman est réobtenu et les hypergraphes dont les graphes associés sont des diagrammes de Hasse de treillis distributifs sont caractérisés.

Introduction

The phylogenetic trees are used to model divisive, bifurcating, filiation, evolutionary, . . . processes. They appear as partially labelled trees. Labelled vertices will be called hereunder *actual vertices* and unlabelled vertices *latent vertices*. For instance, in the context of evolutionary theories actual vertices represent observed individuals (or species) and a latent vertex adjacent to the actual vertices x and y is intended to represent a common unknown ancestor to x and y . In a way latent vertices are what we need to describe a filiation process on observed data.

From a more technical point of view, given a finite set X an X -phylogenetic tree (or shortly an X -tree) is a pair (T, f) of a tree T (with vertex set V and edge

set E) and a map f (the labelling) from X to V such that:

$$\text{if } s \in V - f(X), \text{ then degree of } s \geq 3. \quad (P)$$

Buneman [6] proposes a nice way to reconstruct an X -tree from dichotomies on X . The idea is the following:

The deletion of some edge in T creates two components, hence a bipartition $\sigma = \{A, A'\}$ of X . Such a bipartition is called a *split* of T . The following remarks are straightforward:

1. The splits of T are in one-to-one correspondence with the edges of T .
2. If $\sigma_1 = \{A_1, A'_1\}$ and $\sigma_2 = \{A_2, A'_2\}$ are two distinct splits, then one of the four components A_1, A'_1, A_2 or A'_2 is necessarily contained into an other.
3. The actual vertices may be obtained as the intersection of all the components (of splits) containing them.
4. For each split $\sigma = \{A, A'\}$, a latent vertex v may be located 'on the side' of one component (A or A'). All the components 'containing' v pairwise intersect, but they do not have the Helly property of non-empty total intersection. However, they are sufficient to uniquely determine v .

The Buneman construction deals with the converse problem: given a set B of bipartitions of X , satisfying the property 2 (called *compatibility*) one can, following the steps suggested by 1, 3 and 4, construct an X -tree $T(B)$ whose splits are exactly the elements of B .

Notice that if all the components of the bipartitions fulfil the Helly property, then the X -tree is *labelled* (i.e. $f(X) = V$). This case has been generalized (without any reference to Buneman) by Mulder and Schrijver [9]: a set of bipartitions of X defines a copair hypergraph, i.e. a hypergraph H with vertex set X and hyperedge set \mathcal{E} such that, if $A \in \mathcal{E}$, then $A' \in \mathcal{E}$, with $A' = X - A$. Mulder and Schrijver show that median graphs are equivalent to copair hypergraphs whose hyperedge set fulfils the Helly property (with an additional property of maximality which insures the unicity of the X -label of each vertex of the graph).

Recall that *median graphs* appear as generalizations of both trees and Hasse diagrams of distributive lattices (see Bandelt and Hedlikova [3], Mulder [8]). They are (simple, connected and loopless) graphs such that for each three vertices u, v, w all the shortest paths between these vertices, two by two, have exactly one vertex, called the *median* of u, v, w , in common.

Buneman and Mulder-Schrijver constructions have some points in common, in particular they coincide in the case of labelled trees.

The purpose of this paper is to establish the missing link between these two constructions. Relaxing, both the compatibility and the Helly property, a graph $G(H)$ is associated with each copair hypergraph H . Some vertices of $G(H)$ (the actual vertices) are labelled by the elements of X , others (the latent vertices) are not labelled. Moreover $G(H)$ has the two following properties:

- (i) It is a median graph
- (ii) Each latent vertex may be obtained as an 'iterated median' from actual vertices.

Notice that (ii) appears as a suitable extension of condition (P), which makes each latent vertex the median of three actual vertices.

Graphs fulfilling (i) and (ii) are called X -median graphs. They can be thought, in a way, as graphs modelling filiation processes, when cycles are allowed.

The last part of the paper is devoted to the characterization of copair hypergraphs whose associated X -median graphs belong to some given classes.

Throughout the text, we shall denote by $\#A$ the number of elements of the finite set A .

1. Preliminaries

1.1. Definitions on hypergraphs

A *hypergraph* is a pair $H = (X, \mathcal{E})$, where X is a finite set and \mathcal{E} a set of subsets of X such that \mathcal{E} covers X and \emptyset is not in \mathcal{E} . The elements of X are called the *vertices* of H . The elements of \mathcal{E} are called the *hyperedges* of H . Notice that a hypergraph such that $\#A = 2$, for each $A \in \mathcal{E}$, is nothing but an undirected graph without loops.

With the hypergraph H , its *dual hypergraph* H^* is associated. The vertices of H^* are the hyperedges of H and the hyperedges of H^* are the sets $\mathcal{E}_x = \{A : A \in \mathcal{E} \text{ and } x \in A\}$, for each $x \in X$.

The hypergraph H *separates* X , when, for each $x, y \in X$, there exists $A \in \mathcal{E}$ such that $x \in A$ and $y \notin A$. Clearly $H = H^{**}$ if and only if H separates X (notice that this result becomes false when multiple edges are allowed).

A *clique* in the hypergraph H is a subset \mathcal{F} of \mathcal{E} such that for each $A, B \in \mathcal{F}$, the intersection $A \cap B$ is not empty. The maximal elements of the set of all the cliques of H , ordered by the set theoretic inclusion (between sets of subsets of X) are called the *maximal cliques*. We denote by $C(H)$ the set of all maximal cliques of H .

A clique of H with a nonempty intersection is called a *Helly clique*. H is called a *Helly hypergraph*, when each clique of H is Helly.

The *partial hypergraph* of H induced by a subset \mathcal{F} of the hyperedge set is the hypergraph $H^{\mathcal{F}}$, with \mathcal{F} as hyperedge set and the union of all hyperedges in \mathcal{F} as vertex set. We shall say that H is an *extension* of $H^{\mathcal{F}}$. The *subhypergraph* of H induced by a subset Y of the vertex set X is the hypergraph H_Y , having Y as vertex set and $\mathcal{E}_Y = \{A \cap Y : A \in \mathcal{E}\} - \{\emptyset\}$ as hyperedge set.

1.2. Copair hypergraphs and their maximal cliques

For each subset A of X , we denote by A' its complement. So, $A' = X - A$.

A *copair hypergraph* is a hypergraph $H = (X, \mathcal{E})$, such that $A \in \mathcal{E}$ implies $A' \in \mathcal{E}$. For each $A \in \mathcal{E}$, the pair $\sigma = \{A, A'\}$ is called a *split* of H . We denote by $S(H)$ the set of all the splits of H .

Notice that the number of edges of a copair hypergraph is even. Throughout the paper, we shall assume that $H = (X, \mathcal{E})$ is a copair hypergraph, with $\#\mathcal{E} = 2p$.

Lemma 1. *Let \mathcal{F} be a maximal clique of H . Then, $\#\mathcal{F} = p$.*

Proof. If \mathcal{F} is a clique, then $A \in \mathcal{F}$ implies $A' \notin \mathcal{F}$, so $\#\mathcal{F} \leq p$. Assume that $\#\mathcal{F} < p$, then there exists some hyperedge A so that $A \notin \mathcal{F}$ and $A' \notin \mathcal{F}$. If $\mathcal{F} \cup \{A\}$ is a clique, then \mathcal{F} is not maximal. If not, then there exists $B \in \mathcal{F}$ such that $B \cap A = \emptyset$ and $\mathcal{F} \cup \{A'\}$ is a clique. \square

Notice that for each $x \in X$, \mathcal{E}_x is a maximal (Helly) clique. So the hypergraph $(\mathcal{E}, C(H))$ is an extension of the dual H^* of H . We denote by Φ the mapping from X to $C(H)$ which associates \mathcal{E}_x with x .

A *non-complemented* subset of \mathcal{E} is a subset \mathcal{F} of \mathcal{E} such that, for each $A \in \mathcal{E}$, the assumption that A is in \mathcal{F} implies that A' is not in \mathcal{F} . A non-complemented subset \mathcal{F}_σ is associated with each maximal clique \mathcal{F} and with each split $\sigma = \{A, A'\}$ of H . It is defined as follows: if A (resp. A') $\in \mathcal{F}$, then \mathcal{F}_σ is obtained by replacing A (resp. A') by A' (resp. A) in \mathcal{F} .

Lemma 2. *Let \mathcal{F} be a clique of H and σ be a split of H . The two assertions (i) and (ii) are equivalent:*

(i) \mathcal{F} is a clique.

(ii) *The component of σ situated in \mathcal{F} is a minimal element of \mathcal{F} ordered by inclusion.*

Proof. For $\sigma = \{A, A'\}$ assume that $A \in \mathcal{F}$. Then \mathcal{F}_σ is a clique if and only if the intersection $B \cap A'$ is not empty for each B in $\mathcal{F} - \{A\}$. In other words \mathcal{F}_σ is a clique if and only if there exists no hyperedge B in $\mathcal{F} - \{A\}$ included into A . Hence the result. \square

Lemma 3. *Let \mathcal{F} and \mathcal{G} be two maximal cliques of H . There exists a sequence $\sigma_1, \dots, \sigma_k$ of splits such that:*

(i) $\mathcal{F}_{\sigma_1 \dots \sigma_k} = \mathcal{G}$ and

(ii) $\mathcal{F}_{\sigma_1 \dots \sigma_j}$ is a maximal clique for $j = 1, \dots, k$

Moreover, the smallest value of the integer k is $\#(\mathcal{F} \cap \mathcal{G})$.

Proof. We write $\mathcal{F} = \{A_1, \dots, A_k, A_{k+1}, \dots, A_p\}$, with $\mathcal{F} \cap \mathcal{G} = \{A_{k+1}, \dots, A_p\}$. Denote by σ_1 the split $\{A, A_1'\}$. The result is obtained by induction on k . It is obvious for $k = 1$ (in this case $\mathcal{G} = \mathcal{F}_{\sigma_1}$). Assume $k > 1$. Relabelling the σ_i , we can always imagine that A_1 is minimal in $\{A_1, \dots, A_k\}$, ordered by inclusion. In that case, for each $i \geq 1$, we have that $A_1' \cap A_{k+i}$ is not empty (because \mathcal{G} is a clique). Hence A_1 is minimal in \mathcal{F} and (Lemma 2) \mathcal{F}_{σ_1} is a maximal clique. Now, the induction hypothesis applies to \mathcal{F}_{σ_1} . Hence the result. \square

Lemma 4. *For each split σ there exist maximal cliques \mathcal{F} and \mathcal{G} such that $\mathcal{G} = \mathcal{F}_\sigma$.*

Proof. From Lemma 2, it is sufficient to prove that a hyperedge A is always minimal into some clique. Let \mathcal{F} be a clique; denote by $m(\mathcal{F})$ the number of minimal elements of \mathcal{F} included into A . So, A is minimal in \mathcal{F} if and only if $m(\mathcal{F}) = 0$. Assume $m(\mathcal{F}) > 0$ and consider a minimal element B of \mathcal{F} included into A . We know that \mathcal{F}_τ is a maximal clique, with $\tau = \{B, B'\}$. Moreover $m(\mathcal{F}_\tau) < m(\mathcal{F})$. Hence the result, by induction on $m(\mathcal{F})$. \square

Denote by $\alpha(H)$ the number of maximal cliques of H . Obviously, $\alpha(H) \leq 2^p$. Moreover, it follows from Lemma 3 that $p + 1 \leq \alpha(H)$. Hence:

$$p + 1 \leq \alpha(H) \leq 2^p.$$

We shall establish in Section 3 that these two bounds are attained and the hypergraphs attaining these two bounds will be characterized.

With H is associated the graph $G(H)$ defined as follows:

The vertex set of $G(H)$ is the set $C(H)$ of all maximal cliques of H . Two vertices \mathcal{F} and \mathcal{G} are adjacent in $G(H)$, if and only if $\mathcal{G} = \mathcal{F}_\sigma$, for some split σ .

Proposition 1 hereunder follows from Lemmas 2 and 3.

Proposition 1. *The graph $G(H)$ associated with the copair hypergraph H is connected. The degree of a vertex \mathcal{F} of $G(H)$ is the number of the minimal elements of \mathcal{F} ordered by inclusion.*

Let Φ be the map from X to $C(H)$ such that $\Phi(x) = \mathcal{E}_x$. The mapping Φ allows to represent elements of X as vertices of $G(H)$. Moreover:

- (i) Φ is injective if and only if H separates X ,
- (ii) Φ is surjective if and only if H is a Helly copair hypergraph.

The edges of $G(H)$ represent the splits of H . Consider the mapping μ , from the edge set of $G(H)$ to the set $S(H)$ of all the splits of H , defined by $\mu(\mathcal{F}\mathcal{G}) = \sigma$ if and only if $\mathcal{G} = \mathcal{F}_\sigma$. We know from Lemma 4 that μ is surjective. In case of injectivity, each split is represented by exactly one edge. In the general case define $\beta(\sigma)$ as the number of edges $\mathcal{F}\mathcal{G}$, in $G(H)$, such that $\mu(\mathcal{F}\mathcal{G}) = \sigma$ and define $\beta(H)$ as the sum of all the $\beta(\sigma)$. As the number of vertices of $G(H)$ was $\alpha(H)$, the number of its edges is $\beta(H)$.

The mapping μ is injective if and only if $\beta(H) = p$. This last equality holds if and only if $\alpha(H) = p + 1$ (since $G(H)$ is connected and $\alpha(H) \geq p + 1$). It follows that if μ is one-to-one, then $G(H)$ is a tree. In fact, the converse is also true (Section 3).

Let $\sigma = \{A, A'\}$ be a split of H . Consider the partial hypergraph $H(\sigma) = H^{\mathcal{E}-\sigma}$, induced by $\mathcal{E} - \{A, A'\}$. Let $\beta'(\sigma)$ be the number of maximal cliques \mathcal{F} of H such that \mathcal{F}_σ is not a clique. Clearly: $\beta'(\sigma) + 2\beta(\sigma) = \alpha(H)$ and $\beta'(\sigma) + \beta(\sigma) = \alpha(H(\sigma))$. Hence: $\beta(\sigma) = \alpha(H) - \alpha(H(\sigma))$. From this equality, it is easy to deduce an upperbound for β . It suffices to notice that $\beta(\sigma) > 2^{p-1}$ would imply

$\alpha(H(\sigma)) > 2^{p-1}$ and that we know $\alpha(H(\sigma)) \leq 2^{p-1}$. To summarize, we have:

$$1 \leq \beta(\sigma) \leq 2^{p-1}, \text{ for each split } \sigma,$$

and

$$p \leq \beta(H) \leq p2^{p-1}.$$

1.3. Median graphs

In the following we will need some definitions involving median graphs. For many results about the so-called 'median ternary law', the reader may consult Bandelt and Hedlikova [3].

Let $G = (V, E)$ be an undirected, connected graph without loops. The *interval* $[u, v]$ is the set of all the vertices lying in some shortest path between the vertices u and v .

\mathcal{G} is said to be a *median graph* when, for each $u, v, w \in V$, the intersection $[u, v] \cap [v, w] \cap [u, w]$ contains one and only vertex. This vertex is called the *median* of u, v, w and is denoted by $m(u, v, w)$.

In particular, every hypercube is a median graph. By rooting a hypercube we obtain the Hasse diagram of the boolean lattice of all the subsets of a set K . The graph-theoretic median $m(A, B, C)$ of the three subsets A, B and C of K is just their usual median: $(A \cap B) \cup (B \cap C) \cup (A \cap C)$. This example is essential, since every median graph may be considered as an isometric subgraph of some hypercube closed under the median operation (cf. Mulder [7, 8]).

Two classical examples of median graphs are:

1. Hasse diagrams of distributive lattices. Here, the graph theoretic median coincides with the lattice median (cf. Birkhoff and Kiss [5]): $m(u, v, w) = (u \wedge v) \vee (v \wedge w) \vee (u \wedge w) = (u \vee v) \wedge (v \vee w) \wedge (u \vee w)$.
2. Trees, as illustrated in Fig. 1.

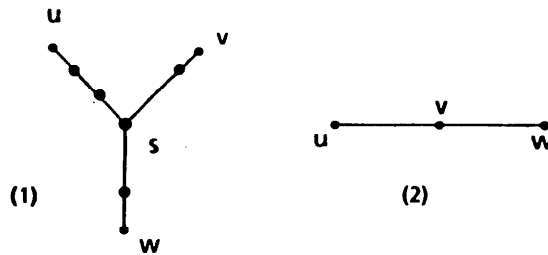


Fig. 1. In (1) $m(u, v, w) = s$. In (2) $m(u, v, w) = v$.

1.4. The Buneman construction and the Mulder-Schrijver theorem revisited

Now, we use the notions and the vocabulary introduced in Section 1.2 to shortly describe the Buneman construction [6] and a part of a result on median graphs obtained by Mulder and Schrijver [9], both emphasized in the introduction.

Buneman considers copair hypergraphs $H = (X, \mathcal{E})$ with the additional property that if A and B are two hyperedges, then one of the four intersections: $A \cap B$, $A \cap B'$, $A' \cap B$, $A' \cap B'$ is empty (Buneman property of compatibility). He establishes that, if H is such a hypergraph, then $G(H)$ is a tree. He interprets the edges of this tree as the splits of H and he distinguishes between two kinds of vertices in $G(H)$:

- the actual vertices representing elements of X (distinct elements, when H separates X),
- the latent vertices, which are not image under Φ of elements of X .

So, actual vertices correspond to maximal cliques with a non-empty intersection and latent vertices correspond to maximal cliques with an empty intersection. Moreover, he constates that every vertex with degree ≤ 2 is an actual vertex.

Mulder and Schrijver establish an equivalence between:

- a Helly copair hypergraph $H = (X, \mathcal{E})$, which separates X , and
- a median graph with vertex set X .

Indeed, this median graph associated with H may be reinterpreted as the graph $G(H)$ defined in Section 1.2.

At the intersection of the results of Buneman and Mulder and Mulder & Schrijver the following would hold:

Let $H = (X, \mathcal{E})$ be a copair hypergraph. The two assertions (i) and (ii) are equivalent:

- (i) $G(H)$ is a tree with X as vertex set
- (ii) H is Helly, separates X and has the Buneman property.

By relaxing the separation and the Helly property, Buneman obtains the trees with latent vertices. By relaxing the Buneman property, Mulder and Schrijver obtain a median graph with X as vertex set (in fact they obtain more, their theorem states an equivalence). In Section 2, by relaxing the separation condition, the Helly property and the Buneman property, we will obtain median graphs with latent vertices. But before doing that, we have to define this last notion.

1.5. Median graphs with latent vertices

Let $G = (V, E)$ be a median graph and consider the sequence (M_n) of mappings from the set of subsets of V into itself defined by: for each Y included in V ,

$$M_0(Y) = Y$$

and

$$M_n(Y) = \{m(u, v, w) : u, v, w \in M_{n-1}(Y)\}.$$

Clearly, $M_n(Y)$ is included in $M_{n+1}(Y)$. The union of all the $M_n(Y)$ is called the *median closure* of Y and denoted by $M(Y)$. When $M(Y) = V$, we say that Y is a *median generator set* of G .

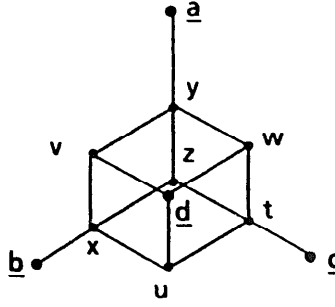


Fig. 2. $X = \{a, b, c, d\}$ and $V = M_2(X)$: $m(a, b, d) = v$, $m(a, b, c) = z$, $m(b, c, d) = u$, $m(a, c, d) = w$. Hence $M_1(X) = \{a, b, c, d, u, v, w, z\}$; $m(b, u, v) = x$, $m(a, v, w) = y$, $m(u, w, c) = t$.

Definition 1. Let X be a set. An X -median graph (X -tree) is a pair (G, Φ) of a median graph (tree) G together with a map Φ from X to V such that $\Phi(X)$ is a median generator set of G .

When Φ is injective, we say that the X -median graph (G, Φ) is *separated*. The smallest integer n such that $M_n(\Phi(X)) = V$ is called the *level* of (G, Φ) . A vertex lying in $\Phi(X)$ is called an *actual vertex*. A vertex not in $\Phi(X)$ is called a *latent vertex*. The following lemma is obvious:

Lemma 5. In an X -median graph (G, Φ) , each vertex with degree ≤ 2 is an actual vertex. Moreover, G is an X -tree, if and only if G is a tree and $\Phi(X)$ contains all the vertices with degree ≤ 2 .

It follows from Lemma 5 that any X -tree is an X -median graph of level 1. Fig. 2 represents an X -median graph of level 2.

An X -median graph (G, Φ) such that Φ is a bijection is called an X -labelled median graph.

2. The structure of the graph $G(H)$

2.1. Structure theorem

In Section 1.1, we have defined the graph $G(H)$ associated with the copair hypergraph $H = (X, \mathcal{E})$. Recall that the vertices of $G(H)$ are the maximal cliques of H and that $\mathcal{F}\mathcal{G}$ is an edge if and only if $\mathcal{G} = \mathcal{F}_\sigma$, for some split σ of H .

In Proposition 1, we established that the graph $G(H)$ is connected. Moreover, the proof gives the form of the shortest paths between two vertices \mathcal{F} and \mathcal{G} : Let A_0, \dots, A_k be the hyperedges in \mathcal{F} but not in \mathcal{G} and consider the splits $\sigma_1 = (A_1, A'_1)$, $1 \leq i \leq k$. The shortest paths between \mathcal{F} and \mathcal{G} are in one-to-one correspondence with the permutations s of $\{\sigma_1, \dots, \sigma_k\}$ such that, for each $j \leq k$, $\mathcal{F}_{s(\sigma_1) \dots s(\sigma_n)}$ is a clique.

In Section 1.2, we have obtained the map Φ from X to the set $C(H)$ of vertices of H . Recall that Φ associates with each $x \in X$ the Helly maximal clique \mathcal{C}_x of all the hyperedges containing x .

Theorem 1. *Let H be a copair hypergraph. Then $(G(H), \Phi)$ is an X -median graph.*

Proof. We have to prove: (i) that $G(H)$ is a median graph and (ii) that the set of all Helly maximal cliques is a median generator set of $G(H)$. The proof will be in three steps. First, we prove (i). Then, the idea, to obtain (ii), is to perform an induction on the number p of hyperedges of H . But, the assumption that, for $p - 1$, each maximal clique of H is in the median closure of the set of all the Helly maximal cliques does not ensure the same result for p . Indeed, Helly maximal cliques, for $p - 1$, come, by restriction, from Helly maximal cliques (for p) and from cliques such that $p - 1$ hyperedges intersect. These cliques will be called ‘almost Helly maximal cliques’ and we shall directly prove, in the second step, that almost Helly maximal cliques are in the median closure of the set of all Helly cliques. Then the third step is devoted to the induction itself.

1. $G(H)$ is a median graph. Consider three maximal cliques \mathcal{F}^1 , \mathcal{F}^2 and \mathcal{F}^3 . We can write:

$$\mathcal{F}^1 = \{A_1, \dots, A_{k-1}, A_k, \dots, A_{q-1}, A_q, \dots, A_{r-1}, A_r, \dots, A_p\},$$

such that, if we denote by \mathcal{A} the set $\{A_1, \dots, A_{k-1}\}$, by \mathcal{B} the set $\{A_k, \dots, A_{q-1}\}$, by \mathcal{C} the set $\{A_q, \dots, A_{r-1}\}$ and by \mathcal{D} the set $\{A_r, \dots, A_p\}$, then:

$$\mathcal{F}^2 = \mathcal{A} \cup \mathcal{B} \cup \mathcal{C}' \cup \mathcal{D}' \quad \text{and} \quad \mathcal{F}^3 = \mathcal{A} \cup \mathcal{B}' \cup \mathcal{C}' \cup \mathcal{D}.$$

where $\mathcal{F}' = \{A', : A \in \mathcal{F}\}$.

Consider a maximal clique \mathcal{F} .

If \mathcal{F} lies on a shortest path between \mathcal{F}^1 and \mathcal{F}^2 , then $\mathcal{A} \cup \mathcal{B} \subseteq \mathcal{F}$.

If \mathcal{F} lies on a shortest path between \mathcal{F}^1 and \mathcal{F}^3 , then $\mathcal{A} \cup \mathcal{D} \subseteq \mathcal{F}$.

If \mathcal{F} lies on a shortest path between \mathcal{F}^2 and \mathcal{F}^3 , then $\mathcal{A} \cup \mathcal{C}' \subseteq \mathcal{F}$.

Now if we examine the intersections in \mathcal{F}^1 , \mathcal{F}^2 and \mathcal{F}^3 , then we see that $\mathcal{F} = \mathcal{A} \cup \mathcal{B} \cup \mathcal{C}' \cup \mathcal{D}$ is a clique. And by construction, \mathcal{F} is the unique element in $[\mathcal{F}^1, \mathcal{F}^2] \cap [\mathcal{F}^3, \mathcal{F}^1] \cap [\mathcal{F}^2, \mathcal{F}^3]$.

2. *The case of almost Helly maximal cliques.* Let $A \in \mathcal{C}$; define an A -almost Helly maximal clique as a maximal clique \mathcal{F} such that: $A \in \mathcal{F}$ and the clique \mathcal{F} has an empty intersection while $\mathcal{F} - \{A\}$ intersects.

Let \mathcal{F} be an A -almost Helly maximal clique. Let σ be the split $\{A, A'\}$. Notice that \mathcal{F}_σ is a Helly clique. Let $\mathcal{F}^0 = \mathcal{F} - \{A\}$. Consider the partition $\mathcal{G}_1, \dots, \mathcal{G}_k$ of \mathcal{F}^0 , where \mathcal{G}_i is a set of hyperedges such that: $G_i \cup \{A\}$ intersects and is maximal for this property in the set $\mathcal{F}^0 - (\mathcal{G}_1, \dots, \mathcal{G}_{i-1})$. So we can find k distinct vertices x_1, \dots, x_k of the hypergraph H such that x_i is both in A and in each hyperedge

of \mathcal{G}_i : $x_1 \in \cap\{G : G \in \mathcal{G}_i\} \cap A$. The maximality property of \mathcal{G}_i ensures that, for j distinct from i , x_1 is not contained in any hyperedge of \mathcal{G}_j . So, if we consider, for each j , $1 \leq j \leq k$, the sets of complements $\mathcal{G}'_j = \{G' : G \in \mathcal{G}_j\}$, then every x_i distinct from x_j lies both in A and in each hyperedge of G'_j . Consider, for each i , with $1 \leq i \leq k$, the set of hyperedges:

$$\mathcal{K}_i = \mathcal{G}'_1 \cup \dots \cup \mathcal{G}'_{i-1} \cup \mathcal{G}_i \cup \mathcal{G}'_{i+1} \cup \dots \cup \mathcal{G}'_k \cup \{A\}.$$

There are exactly p hyperedges in K_i and x_i is in the intersection of all those hyperedges and so K_i is a Helly maximal clique.

Now, observe that it follows from the construction of Step 1, that the median of three maximal cliques is exactly the set of all the hyperedges contained in at least two of them. Applying this principle it is easy to see that \mathcal{F} may be obtained by iterations of the median operation from the K_i and from \mathcal{F}_σ :

$$\mathcal{M}_2 = m(\mathcal{F}_\sigma, \mathcal{K}_1, \mathcal{K}_2) = \mathcal{G}_1 \cup \mathcal{G}_2 \cup \mathcal{G}'_3 \cup \dots \cup \mathcal{G}'_k \cup \{A\},$$

$$\mathcal{M}_3 = m(\mathcal{F}_\sigma, \mathcal{M}_2, \mathcal{K}_3) = \mathcal{G}_1 \cup \mathcal{G}_2 \cup \mathcal{G}_3 \cup \mathcal{G}'_4 \cup \dots \cup \mathcal{G}'_k \cup \{A\},$$

$$\mathcal{M}_{i+1} = m(\mathcal{F}_\sigma, \mathcal{M}_i, \mathcal{K}_{i+1}) = \mathcal{G}_1 \cup \mathcal{G}_2 \cup \dots \cup \mathcal{G}_{i+1} \cup \mathcal{G}'_{i+1} \cup \dots \cup \mathcal{G}'_k \cup \{A\},$$

$$\mathcal{M}_k = m(\mathcal{F}_\sigma, \mathcal{M}_{k-1}, \mathcal{K}_k) = \mathcal{G}_1 \cup \mathcal{G}_2 \cup \dots \cup \mathcal{G}_k \cup \{A\} = \mathcal{F}.$$

3. $(G(H), \varphi)$ is an X -median graph. We proceed by induction on $p = \#\mathcal{E}$. For $p = 1$ or $p = 2$, $G(H)$ admits only Helly cliques and the result is true in that case.

In the general case, consider a maximal clique \mathcal{F} of H , with $\#\mathcal{E} = p$. Let \mathcal{F} be a maximal clique of H and A be a maximal hyperedge in \mathcal{F} . Consider the partial hypergraph H^0 of H induced by $\mathcal{E}^0 = \mathcal{E} - \{A, A'\}$. It follows from Induction hypothesis that $\mathcal{F}^0 = \mathcal{F} - \{A\}$ belongs to the median closure (in $G(H^0)$) of the set of all the Helly cliques of H^0 .

We remark that:

- (i) For each maximal clique \mathcal{G}^0 of H^0 , the set of hyperedges \mathcal{G} , with $\mathcal{G} = \mathcal{G}^0 \cup \{A\}$, is a maximal clique of H . Indeed, for each $B \in \mathcal{G}$, either $B \in \mathcal{F}$ and B intersects A , or else $B' \in \mathcal{F}$ and the assumption that B does not intersect A (i.e. that A is contained in B') would contradict the maximality of A in \mathcal{F} .
- (ii) If \mathcal{G}^0 is the median, in $G(H^0)$, of $\mathcal{G}^0_1, \mathcal{G}^0_2$ and \mathcal{G}^0_3 , then $\mathcal{G}^0 \cup \{A\}$ is the median, in $G(H)$, of $\mathcal{G}_1 \cup \{A\}, \mathcal{G}_2 \cup \{A\}$ and $\mathcal{G}_3 \cup \{A\}$.
- (iii) If \mathcal{G}^0 is a Helly maximal clique of H^0 , then $\mathcal{G}^0 \cup \{A\}$ is either a Helly maximal clique, or else an A -almost Helly maximal clique of H .

From these three remarks, it follows that \mathcal{F} is in the median closure of the set of both Helly and A -almost Helly maximal cliques of H . It is now sufficient to use Step 2 to get the result. \square

2.2. Some remarks

As we have just seen, in the X -median graph $(G(H), \Phi)$, actual vertices are the Helly maximal cliques of H and latent vertices are maximal cliques which are

not Helly. Moreover, every vertex with degree < 3 is an actual vertex (Lemma 5). So vertices with degree > 2 may be constructed from actual vertices with the help of a sequence of ternary median operations. For vertices with degree 3 we get a more precise result:

Proposition 2. *In the graph $G(H)$ associated with the copair hypergraph H each latent vertex with degree 3 is the median of three actual vertices.*

Proof. Let \mathcal{F} be a vertex with degree 3; we know from Proposition 1 that \mathcal{F} , as a maximal clique, admits exactly three minimal hyperedges B , C and D . First notice that if the intersection of B , C and D is not empty, then \mathcal{F} is Helly. So, we can assume that: $B \cap C \cap D = \emptyset$ and there exist three elements t , y and z of X such that t is in $B \cap C$ and not in D ; y is in $B \cap D$ and not in C ; z is in $C \cap D$ and not in B .

Let $\mathcal{N}(x)$ be the set of all the hyperedges, of \mathcal{F} , who do not contain $x \in X$. Clearly \mathcal{F} is the union of $\mathcal{N}(t)$, $\mathcal{N}(y)$ and $\mathcal{N}(z)$. Assume that there exists $A \in \mathcal{N}(t) \cap \mathcal{N}(y)$. Then, t is not in A and both B and C are not included into A . On the other hand, y is not in A , so D is not included in A . This contradicts the fact that B , C and D are exactly the minimal elements of \mathcal{F} . We get $\mathcal{N}(t) \cap \mathcal{N}(y) = \emptyset$ and $\mathcal{N}(t)$, $\mathcal{N}(y)$, $\mathcal{N}(z)$ is a partition of \mathcal{F} . Moreover, $\mathcal{N}(t)' \cup \mathcal{N}(y) \cup \mathcal{N}(z) = \mathcal{E}_t$ and $\mathcal{N}(t) \cup \mathcal{N}(y)' \cup \mathcal{N}(z) = \mathcal{E}_y$ and $\mathcal{N}(t) \cup \mathcal{N}(y) \cup \mathcal{N}(z)' = \mathcal{E}_z$. So \mathcal{F} is the median of the three actual vertices \mathcal{E}_t , \mathcal{E}_y and \mathcal{E}_z . \square

When H is a Helly copair hypergraph, each vertex of $G(H)$ is actual. Hence we get as a corollary of Theorem 1 a part of the Mulder–Schrijver result.

Corollary 1. *Let H be a copair hypergraph. The two assertions hereunder are equivalent:*

- (i) *H separates X and is a Helly hypergraph*
- (ii) *$(G(H), \Phi)$ is an X -labelled median graph.*

We mentioned in Section 1.3 the strong relationship between median graphs and hypercubes. This relationship becomes constructive for the graphs $G(H)$ (hence, according to the Mulder–Schrijver equivalence, for every median graph, assuming H is Helly and separates X): Let $\mathcal{F} = \{A_1, \dots, A_p\}$ be some maximal clique of H . For each maximal clique \mathcal{G} of H , let $L_{\mathcal{F}}(\mathcal{G})$ be the set $\{i: 1 \leq i \leq p \text{ and } A_i' \in \mathcal{G}\}$. The map $L_{\mathcal{F}}$, which associates $L_{\mathcal{F}}(\mathcal{G})$ with \mathcal{G} imbeds the set $C(H)$ of all maximal cliques of H into the set of all the subsets of $\{1, \dots, p\}$. Moreover, if we consider the order on $C(H)$ defined by: $\mathcal{G} \leq \mathcal{H}$ if and only if $L_{\mathcal{F}}(\mathcal{G})$ is included in $L_{\mathcal{F}}(\mathcal{H})$, then the (undirected) Hasse diagram of $(C(H), \leq)$, which appears (via $L_{\mathcal{F}}$) as a subgraph of the p -dimensional hypercube is nothing but the graph $G(H)$. From that remark, one can easily deduce the classical theorem of Avann, which asserts that each median graph is the (undirected) Hasse diagram of some median semilattice ([1], cf. [8]).

The following examples illustrate the interpretation of $G(H)$ (with or without latent vertices) as a piece of a hypercube.

Example 1. (Fig. 3), $X = \{a, b, c, d, e, f, g\}$; \mathcal{E} is defined up to complementation by: $A_1 = \{a, b, c, d, e, f\}$, $A_2 = \{a, b, c, e, g\}$, $A_3 = \{a, b\}$, $A_4 = \{a, c, d\}$.

Example 2. (Fig. 4), $X = \{a, b, c, d\}$, \mathcal{E} is given up to complementation by: $A_1 = \{a\}$, $A_2 = \{a, c, d\}$, $A_3 = \{a, b, d\}$, $A_4 = \{a, b, c\}$, $A_5 = \{a, b\}$.

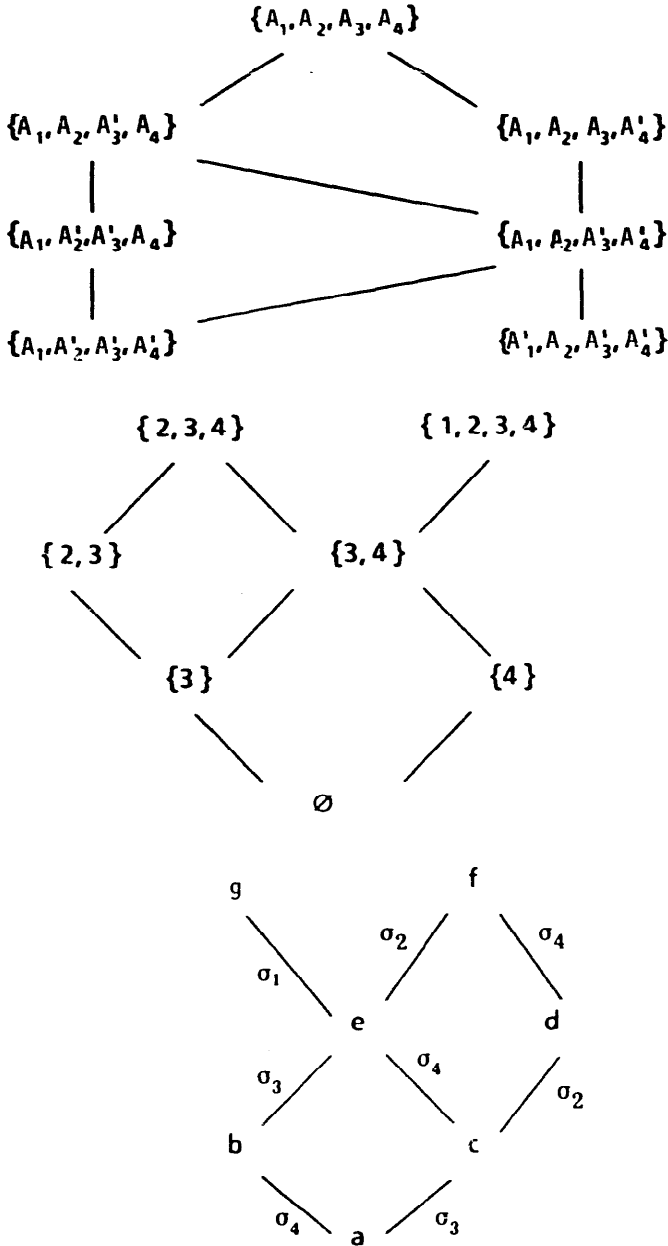


Fig. 3. The maximal cliques, the piece of the cube and the labelled X graph with the map μ .

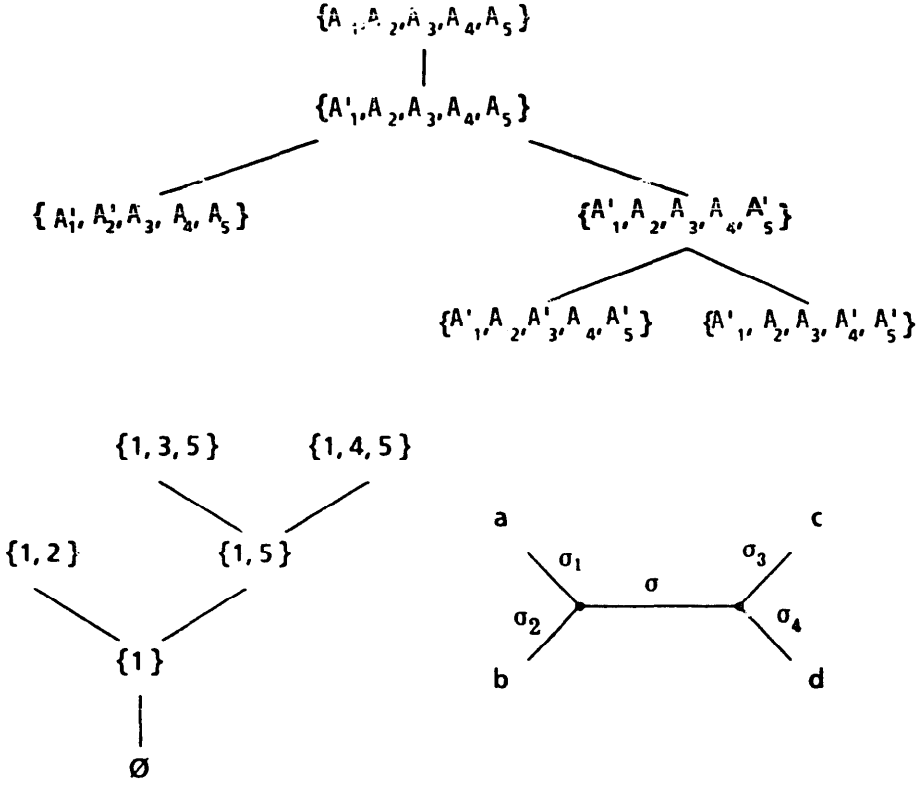


Fig. 4. Maximal cliques, piece of a cube and X-tree.

3. Some special cases

3.1. Special copair hypergraphs

We shall characterize four special structures for $G(H)$ on H , namely: trees, hypercubes, undirected Hasse diagrams of distributive lattices and paths (a path is both a tree and the Hasse diagram of a distributive lattice!). In order to do that, the notions of Buneman hypergraph, cubic hypergraph, distributive hypergraph and Guttman hypergraph are introduced.

Definition 2. Let $\sigma = \{A, A'\}$ and $\tau = \{B, B'\}$ be two distinct splits of the copair hypergraph H . The pair σ, τ is said to be *compatible* if and only if one among the four intersections $A \cap B, A' \cap B, A \cap B', A' \cap B'$ is empty.

Definition 3. A *Buneman hypergraph* is a copair hypergraph such that each pair of distinct splits is compatible.

Definition 4. A *cubic hypergraph* is a copair hypergraph such that no pair of distinct splits is compatible.

In a cubic hypergraph, each non-complemented subset of hyperedges is a clique. And, if \mathcal{F} is a maximal clique, then $\mathcal{F}' = \{A' : A \in \mathcal{F}\}$ is a maximal clique, too. Weakening this property, we get:

Definition 5. A *distributive hypergraph* is a copair hypergraph admitting at least one maximal clique \mathcal{F} such that \mathcal{F}' is a maximal clique.

In Section 3.3 we shall establish that the copair hypergraph H is both Buneman and distributive if and only if the set \mathcal{E} ordered by inclusion admits a maximal chain of length p . Considering H as a 0/1 table (with vertices as rows and hyperedges as columns), we can interpret this situation as a Guttman model (c.f. Barthelemy et al. [4]).

Definition 6. A *Guttman hypergraph* is a copair hypergraph, which is both Buneman and distributive.

3.2. Characterizing Buneman and cubic hypergraphs, with the help of α and β

We have defined, in Section 1.2, the numbers $\alpha(H)$ and $\beta(H)$ associated with the copair hypergraph H . Recall that $\alpha(H)$ is the number of maximal cliques of H (i.e. the number of vertices of $G(H)$). For each split σ , the number of maximal cliques \mathcal{F} such that \mathcal{F}_σ is a clique is denoted by $\beta(\sigma)$ (i.e. $\beta(\sigma)$ is the number of edges, in $G(H)$, ‘representing’ σ) and $\beta(H) = \sum\{\beta(\sigma) : \sigma \in S(H)\}$ (so, $\beta(H)$ is the number of edges of $G(H)$). We have pointed out the following bounds:

$$p + 1 \leq \alpha(H) \leq 2^p,$$

$$1 \leq \beta(\sigma) \leq 2^{p-1},$$

$$p \leq \beta(H) \leq p2^{p-1}.$$

Proposition 3. Let H be a copair hypergraph. Then the four assertions hereunder are equivalent:

- (i) H is a cubic hypergraph
- (ii) $\alpha(H) = 2^p$
- (iii) $\beta(\sigma) = 2^{p-1}$, for each split σ of H
- (iv) $\beta(H) = p2^{p-1}$.

Proof. (i) implies (ii). Clearly, $\alpha(H) = 2^p$ if and only if for each maximal clique \mathcal{F} and for each split σ of H , the set of hyperedges \mathcal{F}_σ is a maximal clique. Now, if H is cubic, consider $\mathcal{F} \in C(H)$ and $A \in \mathcal{F}$. Then, for each $B \in \mathcal{F} - \{A\}$, the hyperedges B and A' intersect. So \mathcal{F}_σ is a maximal clique and $\alpha(H) = 2^p$.

(ii) implies (iii). In Section 1.2 we have the equality: $\alpha(H) - \alpha(H(\sigma)) = \beta(\sigma)$, where $H(\sigma)$ is the partial hypergraph of H induced by $\mathcal{E} - \sigma$. Clearly, if $\alpha(H) = 2^p$, then $\alpha(H(\sigma)) = 2^{p-1}$, hence $\beta(\sigma) = 2^p - 2^{p-1} = 2^{p-1}$.

It is trivial that (iii) implies (iv).

(iv) implies (i). If $\beta(H) = p2^{p-1}$, then $\beta(\sigma) = 2^{p-1}$, for each σ of H . In that case, for each split σ of H , we have: $\alpha(H) = 2^{p-1} + \alpha(H(\sigma))$. By induction on $\alpha(H)$, we see that $\alpha(H) = 2^p$ (this is condition (ii)) and necessarily, H is cubic. \square

Proposition 4. *Let H be a copair hypergraph. Then the four assertions hereunder are equivalent:*

- (i) H is a Buneman hypergraph
- (ii) $\beta(\sigma) = 1$, for each split σ of H
- (iii) $\beta(H) = p$
- (iv) $\alpha(H) = p + 1$.

Proof. Since the equivalence between (ii) and (iii) is trivial, we establish this proposition in two steps.

Step 1: (i) equivalent to (ii).

Step 2: (i) equivalent to (iv).

Step 1. Let H be a Buneman hypergraph. Assume that there exists a split $\sigma = \{A, A'\}$ of H so that $\beta(\sigma) > 1$. Thus we can find two distinct maximal cliques \mathcal{F} and $\mathcal{G} \in C(H)$, such that $A \in \mathcal{F}$ and $A \in \mathcal{G}$, while both \mathcal{F}_σ and \mathcal{G}_σ are maximal cliques. Let $B \in \mathcal{F}$, with $B' \in \mathcal{G}$. The four intersections $A \cap B$, $A' \cap B$, $A \cap B'$, $A' \cap B'$ are not empty. This contradicts the assumption that H is Buneman.

Conversely, we make an induction on p . For $p = 1$, the result is trivial. In the general case we consider a copair hypergraph H , with $\#\mathcal{E} = 2p$ and we assume that for each split σ we have $\beta(\sigma) = 1$. Let M be a maximal element of \mathcal{E} ordered by inclusion. Let $\tau = \{M, M'\}$ and consider the partial hypergraph $H(\tau)$ of H induced by $\mathcal{E} - \tau$. Let β^τ be the index β defined on $H(\tau)$. For each $\pi \in S(H(\tau))$, we have $\beta(\pi) \leq \beta^\tau(\pi)$. Assume that $\beta(\pi) < \beta^\tau(\pi)$. In that case there exists a maximal clique \mathcal{F} of $H(\tau)$ which is not obtained, by restriction from a maximal clique of \mathcal{H} . That is to say that neither $\mathcal{F} \cup \{M\}$, nor $\mathcal{F} \cup \{M'\}$ are maximal cliques of H . So there exist A and B in \mathcal{F} such that $A \subset M \subset B'$. This contradicts the maximality of M . So, we have $\beta^\tau(\pi) = 1$, for each split π of $H(\tau)$ and the induction hypothesis applies to $H(\tau)$: this copair hypergraph is Buneman. It follows that H is Buneman if and only if for each hyperedge A of H , one among the four intersections: $A \cap M$, $A' \cap M$, $A \cap M'$ or $A' \cap M'$, is empty. Notice that either $A' \cap M' = \emptyset$ or $A \cap M' = \emptyset$ would contradict the maximality of M in \mathcal{E} . So, assume there is a hyperedge $A \in \mathcal{E}$, such that M intersects both A and A' . We obtain two cliques $\{A, M\}$ and $\{A', M\}$, each contained in a maximal clique, namely \mathcal{F}^1 and \mathcal{F}^2 . Since $\beta^\tau(\tau) = 1$, either \mathcal{F}_τ^1 or \mathcal{F}_τ^2 is not a clique. So there exists some hyperedge B such that $B \cap M' = \emptyset$. This contradicts the maximality of M . It follows that either $A \cap M' = \emptyset$, or $A' \cap M' = \emptyset$ and H is a Buneman hypergraph.

Step 2. Let σ be a split of H and $H(\sigma)$ be the partial hypergraph of H induced by $\mathcal{F} - \sigma$. We know that $\beta(\sigma) = \alpha(H) - \alpha(H(\sigma))$. If $\alpha(H) = p + 1$, then $\beta(\sigma) = p + 1 - \alpha(H(\sigma))$. Hence $\beta(\sigma) \leq 1$ since $\alpha(H(\sigma)) \geq 1$. Using Step 1, we get the result.

Conversely, assume that H is Buneman and do an induction on p . For $p = 1$, the result is trivial. In the general case, we consider a split σ of H . Obviously $H(\sigma)$ is Buneman and, by induction, $\alpha(H(\sigma)) = p$. Hence the result, since $\alpha(H) = p + \beta(\sigma)$ and $\beta(\sigma) = 1$. \square

In addition, we give a result which specifies the form of each maximal clique in a Buneman hypergraph and provides another characterization of these hypergraphs. In order to do that, let us introduce the following notation: for each hyperedge A of H , we denote by $\mathcal{E}[A]$ the set $\{B \in \mathcal{E} : B' \cap A = \emptyset \text{ or } B' \cap A' = \emptyset\} - \{A'\}$. So $\mathcal{E}[A]$ contains A and all the hyperedges B such that $A \subseteq B$ or $A' \subseteq B$.

Proposition 5. *Let H be a copair hypergraph. Then the three assertions hereunder are equivalent:*

- (i) H is a Buneman hypergraph
- (ii) For each maximal clique \mathcal{F} of H and each hyperedge M minimal in \mathcal{F} , we have: $\mathcal{F} = \mathcal{E}[M]$
- (iii) For each hyperedge A of H , the set $\mathcal{E}[A]$ is a maximal clique.

Proof. (i) implies (ii). In any copair hypergraph H , if \mathcal{F} is a maximal clique of H and if M is minimal in \mathcal{F} , then $\mathcal{F}[M]$ is included into \mathcal{F} . If, moreover, H is Buneman, consider $A \in \mathcal{F}$ such that A' intersects both M and M' . Then either $A \cap M' = \emptyset$ which would contradict the minimality of M , or $A \cap M = \emptyset$ which would contradict the assumption $M \in \mathcal{F}$.

(ii) implies (iii). For each hyperedge A of H , $\mathcal{E}[A]$ is a clique and is contained in a maximal clique \mathcal{F} . If $B \in \mathcal{F}$ is so that $B \cap A' = \emptyset$, then $B' \in \mathcal{E}[A]$, which is impossible. So, A is minimal in \mathcal{F} and $\mathcal{F} = \mathcal{E}[A]$.

(iii) implies (i). Let A and B be hyperedges. Since $\mathcal{E}[A]$ is a maximal clique, either $B \in \mathcal{E}[A]$ or $B' \in \mathcal{E}[A]$. Hence we have the compatibility condition, from the definition of $\mathcal{E}[\sigma]$. \square

3.3. Characterizing distributive hypergraphs and Guttman hypergraphs with the help of the hyperedges ordered by inclusion

Let H be a copair hypergraph, then the examination of the ordered set (\mathcal{E}, \subseteq) is sufficient to decide whenever H is, or is not, a distributive hypergraph.

Proposition 6. *A copair hypergraph H is a distributive hypergraph if and only if the set \mathcal{E} , ordered by inclusion, may be written as the sum of its connected components as follows:*

$$\mathcal{E} = \mathcal{E}_1 + \cdots + \mathcal{E}_k + \mathcal{E}_{k+1} + \cdots + \mathcal{E}_{2k},$$

with $\mathcal{E}_{k+1} = \mathcal{E}'_1 = \{A' : A \in \mathcal{E}_1\}$, for $1 \leq i \leq k$.

Proof. Assume that H is distributive. There exists $\mathcal{F} \in C(H)$, such that $\mathcal{F}' \in C(H)$. Let \mathcal{K} be a connected subset of \mathcal{E} ordered by inclusion, then \mathcal{K} is included into \mathcal{F} or into \mathcal{F}' (if not, we would obtain an inclusion relation between a hyperedge A in \mathcal{F} and a hyperedge B' in \mathcal{F}' , which would imply $A \cap B = \emptyset$ or $A' \cap B' = \emptyset$). Hence, if \mathcal{K} is a connected component, then \mathcal{K}' is a connected component, too and is distinct from \mathcal{K} . The decomposition of \mathcal{E} follows.

Conversely, assume that $\mathcal{E} = \mathcal{E}_1 + \dots + \mathcal{E}_k + \mathcal{E}_{k+1} + \dots + \mathcal{E}_{2k}$ is the decomposition of \mathcal{E} as the sum of its components. In this case $\mathcal{F} = \mathcal{E}_1 + \dots + \mathcal{E}_k$ is a clique of H . For otherwise, there would exist $A \in \mathcal{E}_i$ and $B \in \mathcal{E}_j$ such that $A \cap B = \emptyset$ and neither \mathcal{E}_i nor \mathcal{E}_j would be components. For the same reason, $\mathcal{F}' = \mathcal{E}'_1 + \dots + \mathcal{E}'_k$ is also a clique. Moreover, $\#\mathcal{F}' = p$. Hence \mathcal{F} and \mathcal{F}' are maximal cliques. \square

Corollary 2. *Let H be a copair hypergraph. Then the three assertions hereunder are equivalent:*

- (i) H is a Guttman hypergraph
- (ii) H admits at least one clique which is totally ordered by inclusion
- (iii) The set \mathcal{E} , ordered by inclusion admits a maximal chain of length p .

Proof. (i) implies (ii). If H is distributive, then it admits a maximal clique \mathcal{F} such that \mathcal{F}' is a clique. Moreover, it follows from the proof of Proposition 5 that there is no inclusion relation between hyperedges in \mathcal{F} and hyperedges in \mathcal{F}' . So, if H is Buneman for two hyperedges in \mathcal{F} , the one is always included into the other. Hence \mathcal{F} is a chain.

(ii) implies (iii), that is clear.

(iii) implies (i). If \mathcal{F} is a maximal chain with length p , then \mathcal{F} and \mathcal{F}' are maximal cliques and H is distributive. Moreover if $A \in \mathcal{F}$ (\mathcal{F}'), then for each hyperedge B , either $B \in \mathcal{F}$ and we observe an inclusion between A (A') and B , or else $B \in \mathcal{F}'$ and we observe an inclusion between B' and A (A'). So H is Buneman. \square

Let $\Gamma(H)$ be the number of connected components of the hyperedge set of H , ordered by inclusion. If H is distributive, then $\Gamma(H)$ is even. The converse would be false as indicated below:

Example 3: (Fig. 5). $X = \{a, b, c, d, e, f, g, h, i, j\}$, \mathcal{E} is given, up to complementation, by: $A_1 = \{a, d, i\}$, $A_2 = \{b, h, f\}$, $A_3 = \{c, e, g, j\}$, $A_4 = \{a, e, f, g\}$, $A_5 = \{c, d, h\}$, $A_6 = \{b, i, j\}$.

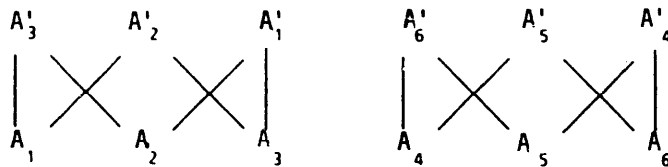


Fig. 5.

Generally, we remark that, if \mathcal{K} is a connected component of \mathcal{E} , then either $\mathcal{K} \cap \mathcal{K}' = \emptyset$ and \mathcal{K}' is a component too, or else $\mathcal{K} = \mathcal{K}'$. Hence the general decomposition of the hyperedge set of a copair hypergraph is: $\mathcal{E} = \mathcal{E}_1 + \dots + \mathcal{E}_k$, with $\mathcal{E}_i = \mathcal{E}'_i$ or $\mathcal{E}_i = \mathcal{E}'_j$, for some j distinct from i . Observe that in Example 3, we had $\mathcal{E} = \mathcal{E}_1 + \mathcal{E}_2$, with $\mathcal{E}_1 = \mathcal{E}'_1$ and $\mathcal{E}_2 = \mathcal{E}'_2$.

Corollary 3. *If H is a Buneman hypergraph, then $\Gamma(H) \leq 2$ and moreover $\Gamma(H) = 2$ if and only if H is Guttman.*

Proof. Let $\mathcal{E} = \mathcal{E}_1 + \dots + \mathcal{E}_k$ be the decomposition of \mathcal{E} as the sum of its components. If $k = 1$, we get the result. Suppose $k > 1$. Let \mathcal{F} be a maximal clique of H . Consider $A \in \mathcal{F} \cap \mathcal{E}_i$ and $B \in \mathcal{F} \cap \mathcal{E}_j$, for some i, j , with $1 \leq i < j \leq k$. Since H is Buneman and \mathcal{E}_i and \mathcal{E}_j are components, we get either $A \cap B = \emptyset$ or $A' \cap B' = \emptyset$. So, $\mathcal{E}_i = \mathcal{E}'_j$ and we get: $\mathcal{E} = \mathcal{E}_1 + \dots + \mathcal{E}_q + \mathcal{E}'_1 + \dots + \mathcal{E}'_q + \mathcal{R}$, where the \mathcal{E}_i and \mathcal{E}'_i are components, $\mathcal{R} \cap \mathcal{F} = \emptyset$, and $\mathcal{R} = \mathcal{R}'$. But necessarily \mathcal{R} is empty (if $A \in \mathcal{R}$, then $A' \in \mathcal{R}$, while A or else $A' \in \mathcal{F}$, this contradicts $\mathcal{R} \cap \mathcal{F} = \emptyset$). \square

3.4. The structure of the graph $G(H)$ in some special cases

Theorem 2. *Let H be a copair hypergraph. Then:*

- (i) $G(H)$ is a tree if and only if H is a Buneman hypergraph
- (ii) $G(H)$ is the (undirected) Hasse diagram of some distributive lattice if and only if H is a distributive hypergraph
- (iii) $G(H)$ is a path if and only if H is a Guttman hypergraph
- (iv) $G(H)$ is a cube if and only if H is a cubic hypergraph.

Proof. (i) If H is Buneman, then according to the results of Section 3.2, $G(H)$ admits $p + 1$ vertices and p edges. Hence it is a tree. Conversely, assume that $G(H)$ is a tree and suppose that the splits $\sigma = \{A, A'\}$ and $\tau = \{B, B'\}$ are not compatible. We get four maximal cliques $\mathcal{F}^1, \mathcal{F}^2, \mathcal{F}^3$ and \mathcal{F}^4 such that:

$$A \in \mathcal{F}^1, B \in \mathcal{F}^1; A \in \mathcal{F}^2, B' \in \mathcal{F}^2; A' \in \mathcal{F}^3, B \in \mathcal{F}^3 \quad \text{and} \quad A' \in \mathcal{F}^4, B' \in \mathcal{F}^4.$$

Consider the paths c_1, c_2, c_3 and c_4 between \mathcal{F}^1 and \mathcal{F}^2 , \mathcal{F}^2 and \mathcal{F}^3 , \mathcal{F}^3 and \mathcal{F}^4 , \mathcal{F}^1 and \mathcal{F}^4 respectively. Clearly:

A is contained in every vertex of c_1 ;

B' is contained in every vertex of c_2 ;

A' is contained in every vertex of c_3 ;

B is contained in every vertex of c_4 .

It follows that two \mathcal{F}^i who are not the extremities of the same c_j are not located on that c_j . So, c_1, c_2, c_3, c_4 constitute a cycle and $G(H)$ is not a tree.

(ii) Assume that H is distributive. Let \mathcal{F} be a maximal clique such that \mathcal{F}' is a clique, too. Consider the map $L_{\mathcal{F}}$ defined as in Section 2.3. This map induces an order relation on $C(H)$: $\mathcal{G} \leq \mathcal{H}$ if and only if $L_{\mathcal{F}}(\mathcal{G})$ is included into $L_{\mathcal{F}}(\mathcal{H})$ and clearly $G(H)$ is the Hasse diagram of $(C(H), \leq)$. Moreover, since $G(H)$ is a median graph, we know from Bandelt [2] that $(C(H), \leq)$ is a median semilattice as a poset. A median semilattice with a greatest element is a distributive lattice. Hence the result, since \mathcal{F}' is the greatest element of $(C(H), \leq)$.

Conversely, assume that $G(H)$ is the undirected Hasse diagram of some distributive lattice. Let \mathcal{F} be the smallest element of that lattice. Going up from \mathcal{F} , we obtain that $\mathcal{G} \leq \mathcal{H}$ if and only if $L_{\mathcal{F}}(\mathcal{G})$ is included into $L_{\mathcal{F}}(\mathcal{H})$. Let \mathcal{Q} be the greatest element of this lattice. Assume that $\mathcal{F} = \{A_1, \dots, A_p\}$, so we can write: $\mathcal{Q} = \{A'_1, \dots, A'_k, A_{k+1}, \dots, A_p\}$. For $k = p$, we have $\mathcal{Q} = \mathcal{F}'$. Assume $k < p$. Then there exists either $j \leq k$, such that: $A'_{k+1} \cap A'_j = \emptyset$; or else $i > k + 1$, such that: $A'_{k+1} \cap A_i = \emptyset$. In the first case $\{A_i, A_{k+1}\}$ is contained into a maximal clique incomparable with \mathcal{Q} (with respect to the lattice order). This is impossible, since \mathcal{Q} is the greatest element of the lattice. In the second case, we get a similar conclusion with $\{A'_{k+1}, A'_i\}$. Hence $k = p$ and $\mathcal{Q} = \mathcal{F}'$, and so hypergraph H is distributive.

(iii) comes from the definition of a Guttman hypergraph and from (i) and (ii).

(iv) When H is cubic, then $L_{\mathcal{F}}$ is an order isomorphism. Hence the result. \square

The part (i) of Theorem 2 is nothing but the Buneman construction (cf. Section 1.4). Looking at the proof of part (ii), we find that only the assumption that $(C(H), \leq)$ has a greatest element works. So we can state:

Corollary 4. *Let H be a copair hypergraph. Then the four assertions hereunder are equivalent:*

- (i) H is a distributive hypergraph
- (ii) $G(H)$ is the undirected Hasse diagram of a poset with both a greatest and a smallest element
- (iii) $G(H)$ is the undirected Hasse diagram of some lattice
- (iv) $G(H)$ is the undirected Hasse diagram of some distributive lattice.

Acknowledgement

The author wishes to thank the anonymous referees for valuable help.

References

- [1] S.P. Avann, Metric ternary distributive semilattices, Proc. Amer. Math. Soc. 12 (1961) 407–414.
- [2] H.J. Bandelt, Discrete ordered sets whose covering graphs are median, Proc. Amer. Math. Soc. 91 (1984) 6–8.

- [3] H.J. Bandelt and J. Hedlikova, Median algebra, *Discr. Math.* 45 (1983) 1–30.
- [4] J.P. Barthelemy, C. Flament and B. Monjardet, Ordered sets in social sciences, in: I. Rival (Ed.) *Ordered sets* (Reidel, Dordrecht, 1982) 721–758.
- [5] G. Birkhoff and S.A. Kiss, a ternary operation in distributive lattices, *Bull. Amer. Math. Soc.* 53 (1947) 749–752.
- [6] P. Buneman, The recovery of trees from measures of dissimilarity, in: F.R. Hodson, D.G. Kendall and P. Tautu (Eds). *Mathematics in historical and archeological sciences* (Edinburgh University Press, Edinburgh, 1971) 387–395.
- [7] H.M. Mulder, n -Cubes and median graphs, *J. Graph Theory* 4 (1980) 107–110.
- [8] H.M. Mulder, The interval function of a graph, *Math. Centre Tract* 132 (Mathematisch Centrum, Amsterdam, 1980).
- [9] H.M. Mulder and A. Schrijver, Median graphs and Helly hypergraphs, *Discr. Math.* 25 (1979) 41–50.